

# **Conséquences génétiques d'un épisode d'adaptation à l'hôte chez l'agent de la rouille foliaire du peuplier**

Antoine Persoons  
Fabien Halkett  
Sébastien Duplessis  
Stéphane De Mita

Séminaire CBGP Montpellier – 28 juin 2016

# Interactions hôtes-pathogènes : un moteur de l'évolution



- **Hypothèse de la Reine rouge :**

Probabilités d'extinction  
indépendantes de l'âge des taxons

Mais liées à l'écosystème

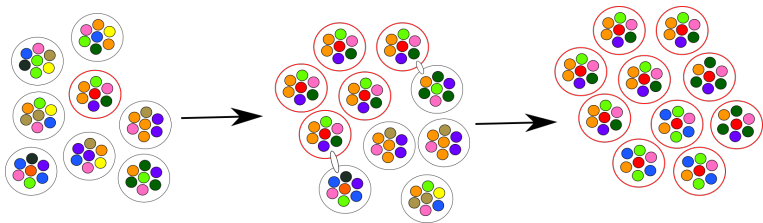
Van Valen *Evol. Theor.* **1** (1973)

- Adaptation constante à  
l'environnement

Pas d'espèces plus adaptées que  
d'autres

- Interactions biotiques imposent une  
coévolution constante

# Conséquences démographiques et génétiques des épisodes d'adaptation



- Interactions hôtes-pathogènes  
→ sélection intense et répétée
- Sélection aux locus adaptatifs  
→ balayage sélectifs dans les régions liées  
→ goulots d'étranglement à l'échelle des populations
- Influence du régime de reproduction / structure de population  
(probabilité de recombinaison)

# *Melampsora larici-populina*



## *Melampsora larici-populina*

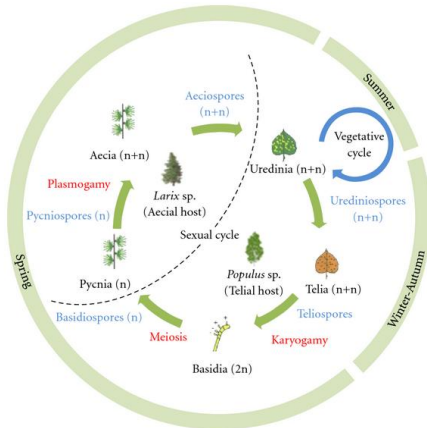
- Royaume : Champignons
- Phylum : Basidiomycota
- Subphylum : Pucciniomycotina
- Classe : Pucciniomycètes
- Ordre : Pucciniales (rouilles)
- Famille : Melampsoraceae

## Pathogène biotrophe obligatoire

- → peupliers sauvages
- → peupleraies cultivées  
(hybrides peu résistants)



# Cycle de vie de *M. larici-populina*



## Cycle hétéroïque et macrocyclique

Figure : Hacquard et al. *J. Pathogens* **2011** (2011)

# Caractéristique du cycle de vie



## Sur peuplier

- De mars à septembre
- Reproduction strictement clonale
- Pas de recombinaison
- Forte croissance (si milieu favorable)
- Dispersion importante



## Sur mélèze

- Nécessaire pour compléter le cycle
- Une méiose / fécondation
- **Dispersion pour trouver l'hôte**

# Répartition des hôtes de *M. larici-populina*

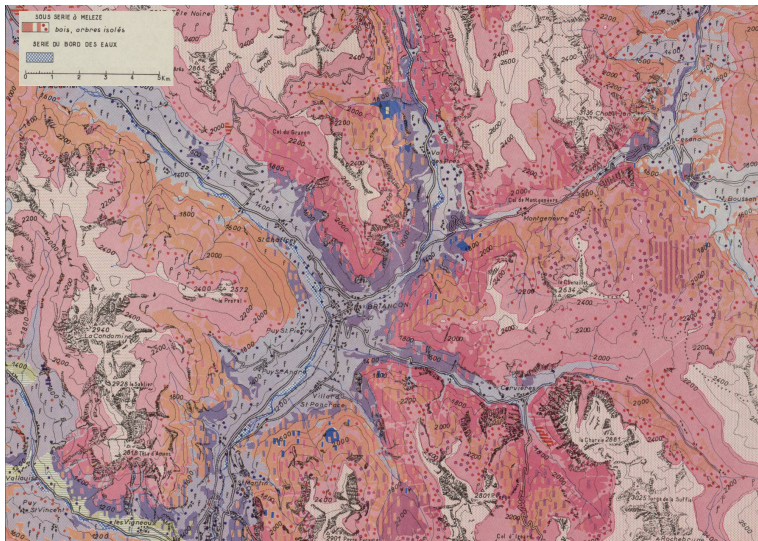


Mélèze d'Europe : Alpes internes en altitude

Peupliers noir : commun (ripisylves et milieux humides)

*Larix* spp. et *Populus* × spp. plantés partout en plaine

# Répartition des hôtes de *M. larici-populina*



# Conséquence évolutives du cycle de vie

## **Deux hôtes différents non sympatriques mais en contact**

- Dans l'aire d'origine : contact localisé
- Plantations de peupliers
- Mélèze ornemental et plantations

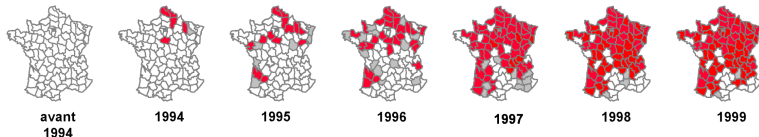
## **Faible consanguinité**

- Dispersion par le vent
- Reproduction sexuée annuelle imposée
- Types sexuels
- Nécessité de disperser pour changer d'hôte

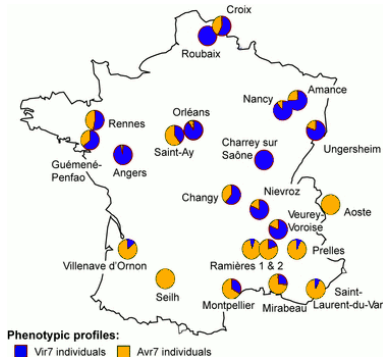
# Contournement des résistances de l'hôte

- Cultivars commerciaux / résistances qualitatives (totales)
- Probablement monogéniques
- Successivement contournées par *M. larici-populina*
  - 1986 Contournement R3 (cv. Luis-Avanzo)
  - 1994 Contournement R7 (cv. Beauprê)
  - 1997 Contournement R8 (cv. Hoogvorst)
- Faibles résistances quantitatives
- → Dommages importants

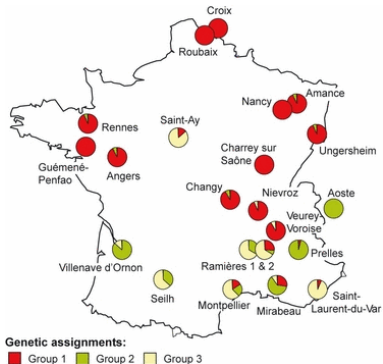
Propagation rapide du contournement de R7 :



# Résultat du contournement de R7



Répartition des individus Vir7



Clusters génétiques (SSR)

- Individus Vir7 / groupe 1 → groupe cultivé virulent
- individus Avr7 / groupe 2 → sauvages (moins de plantations dans le Sud)

# Le contournement de résistance par *Melampsora larici-populina*, l'agent de la rouille du peuplier : impact démographique et déterminisme génétique

Thèse Antoine Persoons



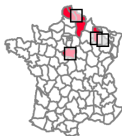
- Analyse du génome de *M. larici-populina*  
Persoons et al. *Frontiers Plant Sci.* **5** (2014)
- Structure des populations (532 isolats, 21 microsatellites)
- Modèle historique et détection de la sélection (86 isolats, > 1M SNP)



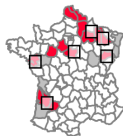
# Échantillonnage historique



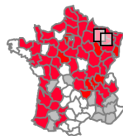
avant  
1994



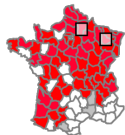
1994



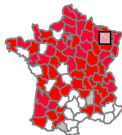
1995



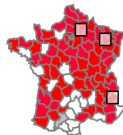
1997



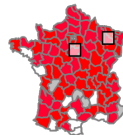
1998



1999



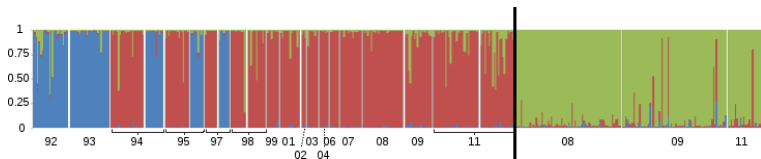
2009



2011

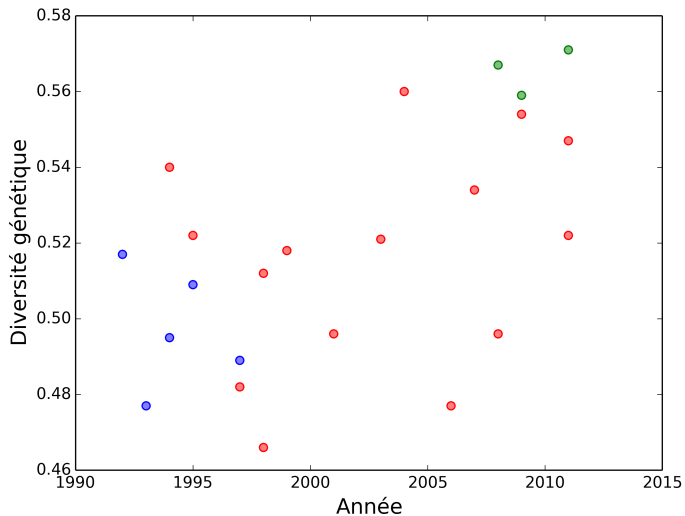
- Collection historique maintenue au laboratoire
- Souches maintenues par reproduction asexuée
- >600 isolats sélectionnés (1992-2011)
- 25 marqueurs SSR

# Analyse de la structure



- Structure optimale à  $K = 3$  clusters (TESS)
- Groupe ancestral **Avr7** éteint vers 1998
- Remplacement par le groupe **Vir7**
- Groupe **sauvage** dans le Briançonnais

# Augmentation de la diversité au sein des populations Vir7



# Approche par reséquençage du génome

## Sous-échantillonnage

- 21 individus **Avr7** (1993)
- 22 individus **Vir7** (1994)
- 22 individus **Vir7** (1998)
- 21 individus **sauvages** (2008)

## Génome de référence : 98AG31 (groupe **Vir7** 1998)

- 101 Mbp
- 45 % ADN répété
- Assemblage 18 chromosomes (max. 8.6 Mbp)
- N50 : 5.2 Mbp
- Annotation encore incomplète

## Séquençage Illumina

- Paired-end, 2 × 150 bp, 69–116 X

# Détection du polymorphisme

## Étude préliminaire (14 échantillons, 22–46 X)

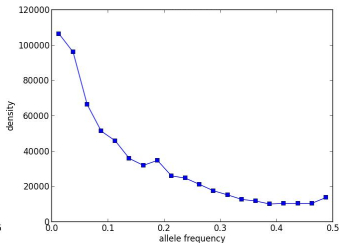
→ 367 412 SNP

Persoons *et al.* *Frontiers Plant Sci.* **5** (2014)

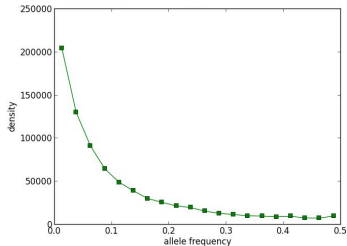
## Cette étude (86 échantillons)

- 25–143 millions de lectures / individu (moyenne 68 millions)
- Alignement *bwa* (max. 2 mismatches)
- 78 % lectures alignées → profondeur finale 32–158 X (moyenne 79 X)
- → 1 109 646 SNP et 53 809 indels ( $\pm 1$  SNP / 100 pb)
- Filtre : MAF (allèle minoritaire) minimum 0.05
- → 528 517 SNP conservés
- Nombre de sites considérés : 96 104 152
- $\pi$  par site : 0.0015

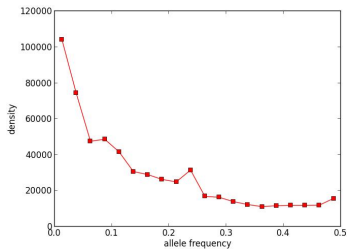
# Spectre de fréquences par population



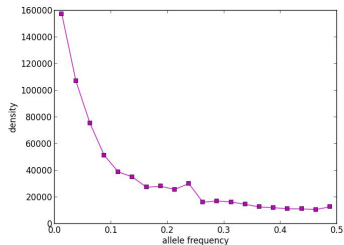
**Avr7 1993**



**Sauvages 2008**

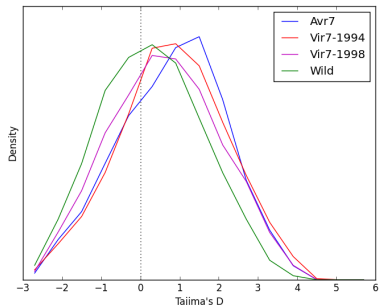
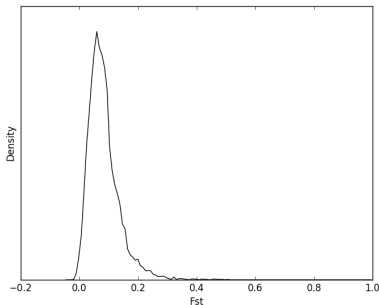


**Vir7 1994**



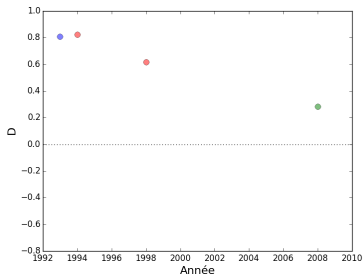
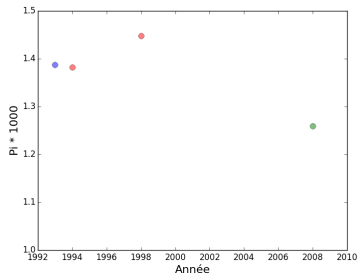
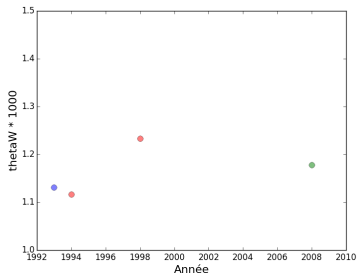
**Vir7 1998**

# Structure du polymorphisme



- Niveau de diversité conséquent (1 SNP pour 100 bp avant filtre)
- Populations différenciées ( $F_{ST} = 0.085$ )
  - Vir7 1994 vs Vir7 1998 :  $F_{ST} = 0.036$  - les plus proches
  - Vir7 1994 vs sauvages :  $F_{ST} = 0.084$  - un peu + faible
  - Vir7 1994 vs Avr7 :  $F_{ST} = 0.104$
  - Avr7 vs sauvages :  $F_{ST} = 0.102$
  - Vir7 1998 vs Avr7 :  $F_{ST} = 0.131$
  - Vir7 1998 vs sauvages :  $F_{ST} = 0.114$  -  $F_{ST}$  en augmentation

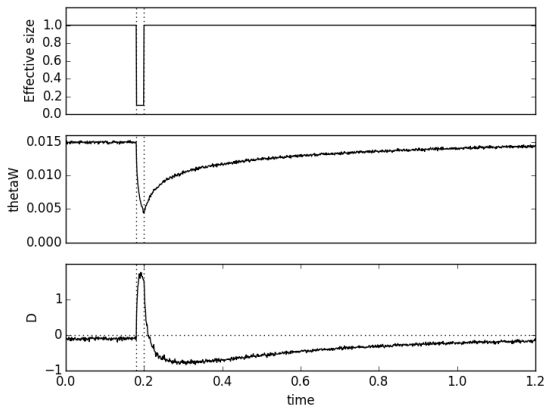
# Augmentation de la diversité chez les Vir7





# Déséquilibre des fréquences alléliques

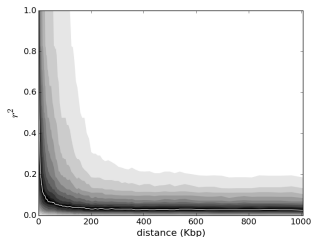
- Sous-structure probable dans les échantillons cultivés
- Goulots d'étranglement peuvent déformer le spectre de fréquences



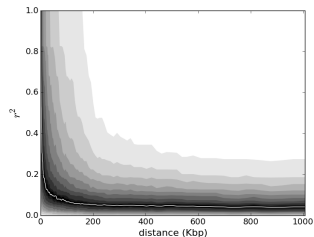
Paramètres de simulation :

- $n = 100$  individus haploïdes
- $\theta = 15$   
(0.015 par site)
- bottleneck  
 $0.02 \times 4N$  gén.  
 $N_b = 0.05 \times N_0$

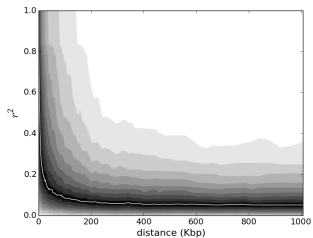
# Déséquilibre de liaison



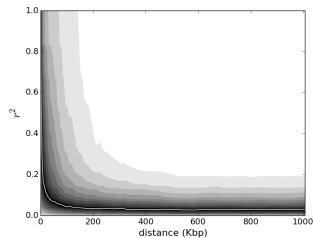
**Sauvages 2008**



**Avr7 1993**



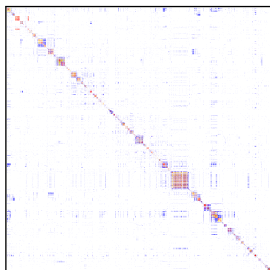
**Vir7 1994**



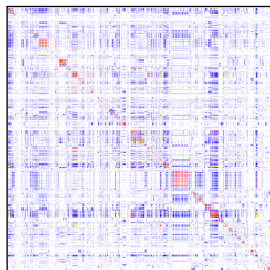
**Vir7 1998**

# Déséquilibre de liaison

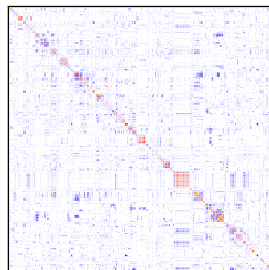
$r^2$  entre SNPs du chromosome 1



**Sauvages 2008**



**Vir7 1994**



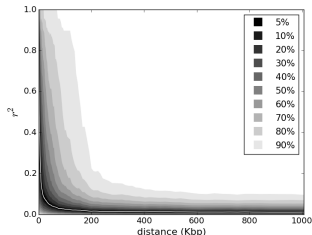
**Vir7 1998**

Déséquilibre "de fond" non lié à la distance physique  
particulièrement immédiatement après le contournement  
→ effet de fondation

# Inférence démographique (préliminaire)

## Données utilisées

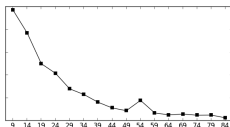
- Fenêtres (2000 bp) le long du génome
- Distance minimum 200 Kbp
- $\geq 1$  SNP par fenêtre



- → **476 fenêtres**
- Max 243 SNP / fenêtre (moyenne : 21)
- Total 10 014 SNP (MAF  $\geq 0.05$ )

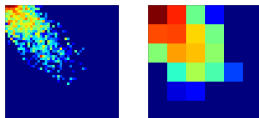
# Statistiques résumantes

- Spectre de fréquences (16 catégories)



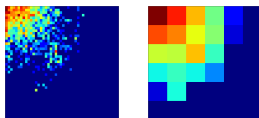
- Spectre conjoint (6 × 6 catégories)

Vir7 (1994) vs Vir7 (1998)



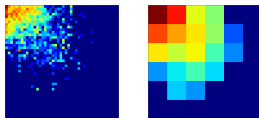
- Spectre conjoint (6 × 6 catégories)

sauvages vs Vir7 (1994)



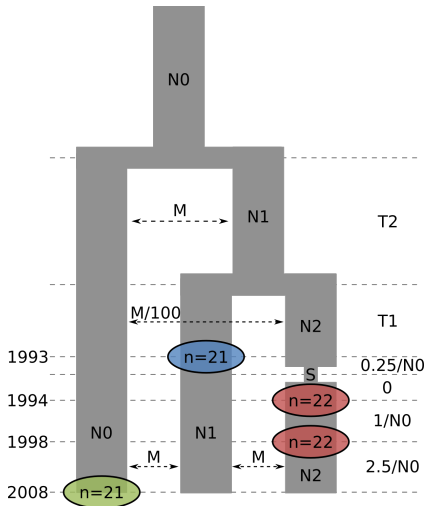
- Spectre conjoint (6 × 6 catégories)

sauvages vs Avr7



Total :  $16 + 6 * 6 * 3 = 124$  statistiques

# Modèle démographique



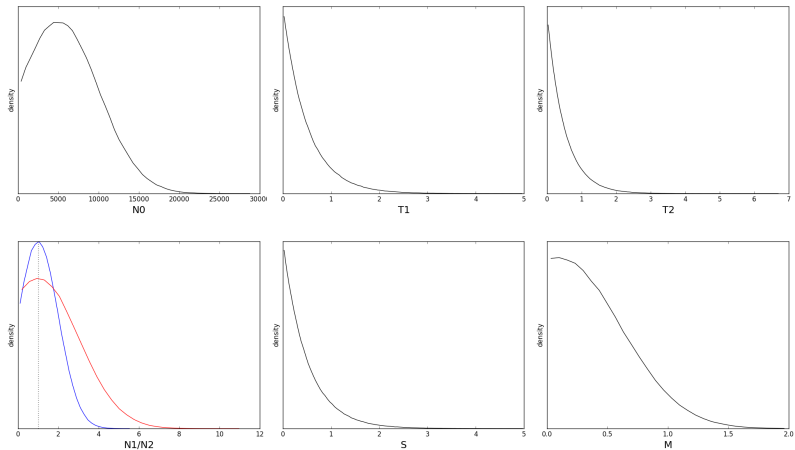
## 7 paramètres :

- **$N_0$**  – pop sauvage (individus)
- **$N_1$**  – pop avirulente (relative)
- **$N_2$**  – pop virulente (relative)
- **$M$**  – taux de migration
- **$S$**  – intensité du bottleneck
- **$T_1$**  – temps / fusion pop virulente
- **$T_2$**  – temps / fusion finale

## 2 modèles :

- **$M_1$**  – tel que représenté
- **$M_2$**  – fusion de la pop virulente avec les sauvages en premier

# Priors



# Ajustement par ABC

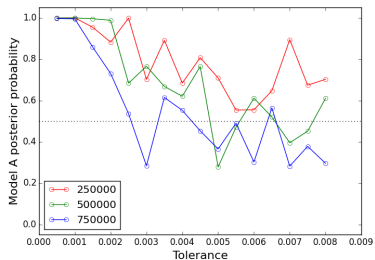
- Simulations
  - 476 fragments non recombinants, non liés entre eux
  - Nombre de mutations fixé (1 - 243 SNP / fenêtre)
  - Minimum MAF = 0.05
  - → 1 435 000 (M1) et 1 688 000 (M2) simulations
- Calcul des 124 statistiques
- Choix de modèle
  - Réseau neuronal
  - Tolérance :  $5 \cdot 10^{-4}$  -  $8 \cdot 10^{-3}$  16 valeurs
  - Sous-ensemble des données (250 000, 500 000, 750 000)
- Ajustement
  - Réseau neuronal
  - Tolérance :  $1.5^{-3}$  (> 2000 simulations acceptées)

Analyse ABC avec le package `abc` sous R

Calcul statistiques et simulations avec `EggLib` sous Python

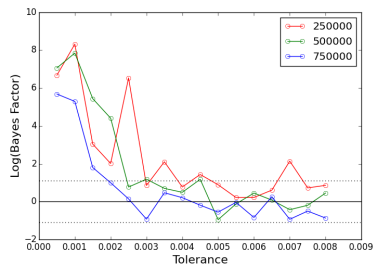


# Choix de modèle

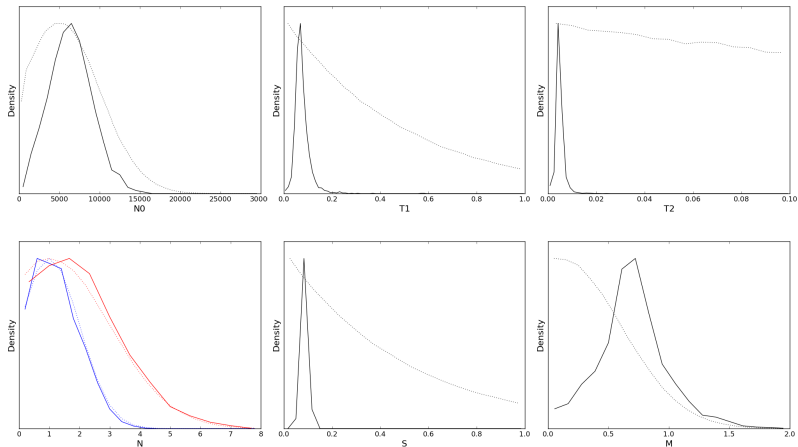


Le modèle M1  
semble meilleur

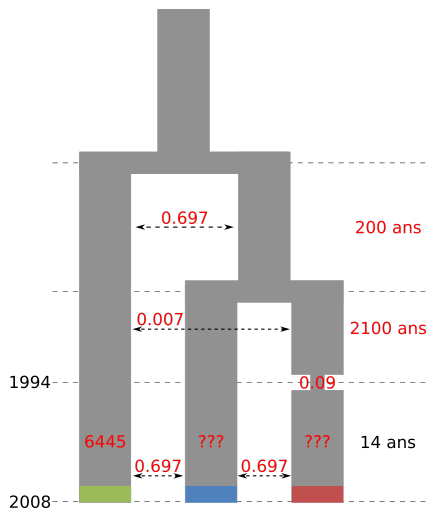
(fusion entre pop **Vir7** et **Avr7** en premier)



# Posteriors



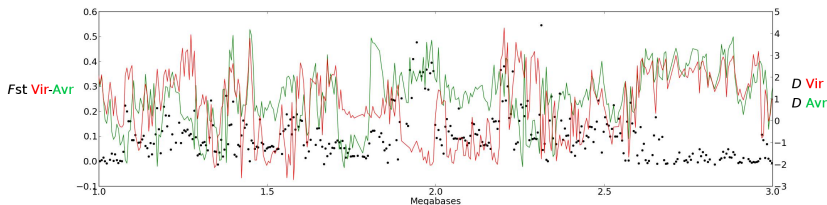
# Modèle avec paramètres ajustés



## Biais possibles :

- 2 modèles considérés sur 3 possibles  
population **Vir7** ancestrale ?
- Sous-structuration ignorée
- Rapport des taux de migration fixé
- Pas de variance de migration efficace

# Recherche des régions sous sélection positive



- Fenêtres de 5000 bp
- Simulations postérieures conditionnées au modèle
- Recherche de fenêtres avec forts  $F_{ST}$  et  $|D|$  **Vir7**  
→ 20 fenêtres (14 gènes)
- $D$  **Vir7** négatif dans deux régions du LG15  
→ Deux gènes "unknown function"
- Patron en **îlots de différenciation**

# Conclusions

- Utilité des données temporelles
- Contournement de résistance → remplacement de population
- Pas d'excellent candidat
  - Type de mutation non considéré (délétion)?
  - *Soft sweep*?
  - Plusieurs gènes?
- Goulot d'étranglement **modéré**
  - Flux de gènes intenses?
  - Origine plus ancienne mais non détecté?
  - Immigration (*soft sweep*)?

Antoine Persoons



Fabien Halkett



Sébastien Duplessis



Axelle Andrieux



Bénédicte Fabre



INRA Nancy UMR Interactions arbres-microorganismes